(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 February 2008 (14.02.2008)

PCT

(10) International Publication Number
WO 2008/018887 A1

(51) International Patent Classification:
G06K 9/00 (2006.01) *G06K 9/34* (2006.01)
G06K 9/32 (2006.01) *G06K 9/40* (2006.01)

(21) International Application Number:
PCT/US2006/032959

(22) International Filing Date: 22 August 2006 (22.08.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/464,083 11 August 2006 (11.08.2006) US

(71) Applicant *(for all designated States except US)*: **FOTONATION VISION LIMITED** [IE/IE]; Galway Business Park, Dangan, Galway City, County Galway (IE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CORCORAN, Peter** [IE/IE]; Cregg, Claregalway, Co. Galway (IE). **STEINBERG, Eran** [US/US]; 137 Granville Way, San Francisco, CA 94127 (US). **PETRESCU, Stefan** [RO/RO]; Str. Cap. Gheorghe Ion Nr. 4, B1.73, Sc.3, Et.2, Ap.37, Sector 4, Bucharest (RO). **DRIMBAREAN, Alexandru** [RO/RO]; 26 Armata Romana, 2200 Brasov (RO). **NANU, Florin** [RO/RO]; Str. Izvorul Muresului, NR13, BI D22, Sc D, Ap 32, Sector 4, Bucuresti (RO). **POSOSIN, Alexei** [MD/RO]; 28 Arus na Tra, Cois Cuain Dr., Colohan Rd., Galway (IE). **BIGIOI, Petronelnel** [RO/RO]; 57 Scelig Ard, Headford Road, Galway, Co. Galway (IE).

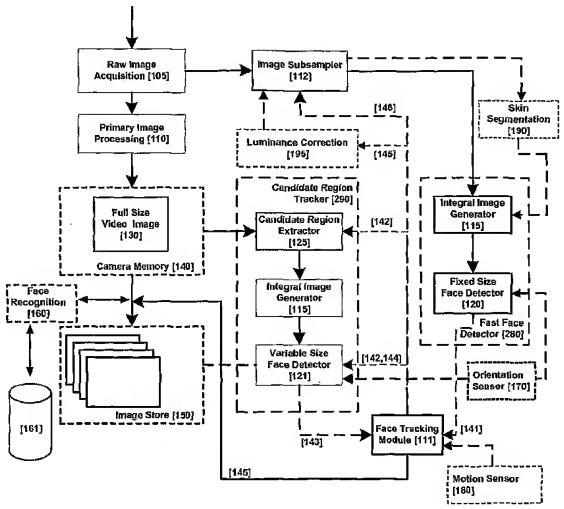
(74) Agent: **SMITH, Andrew, V.**; JACKSON & CO., LLP, 6114 La Salle Avenue, #507, Oakland, CA 94611-2802 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

[Continued on next page]

(54) Title: REAL-TIME FACE TRACKING IN A DIGITAL IMAGE ACQUISITION DEVICE



(57) Abstract: An image processing apparatus for tracking faces in an image stream iteratively receives a new acquired image from the image stream, the image potentially including one or more face regions. The acquired image is sub-sampled (1 12) at a specified resolution to provide a sub-sampled image. An integral image is then calculated for a least a portion of the sub-sampled acquired image. Fixed size face detection (120) is applied to at least a portion of the integral image to provide a set of candidate face regions. Responsive to the set of candidate face regions produced and any previously detected candidate face regions, the resolution at which a next acquired image is sub-sampled is adjusted.

WO 2008/018887 A1

European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

REAL-TIME FACE TRACKING IN A DIGITAL IMAGE ACQUISITION DEVICE

PRIORITY

This application claims priority to United States patent application no. 11/464,083, filed August 11, 2006, and is hereby incorporated by reference.

Field of the Invention

The present invention provides an improved method and apparatus for image processing in acquisition devices. In particular the invention provides improved real-time face tracking in a digital image acquisition device.

Background of the Invention

Face tracking for digital image acquisition devices describe methods of marking human faces in a series of images such as a video stream or a camera preview. Face tracking can be used for indication to the photographer the locations of faces in an image, improving the acquisition parameters, or for allowing post processing of the images based on knowledge of the location of faces.

In general, face tracking systems employ two principle modules: (i) a detection module for location of new candidate face regions in an acquired image or a sequence of images; and (ii) a tracking module for confirmed face regions.

A well-known fast-face detection algorithm is disclosed in US 2002/0102024, Viola-Jones. In brief, Viola-Jones first derives an integral image from an acquired image – usually an image frame in a video stream. Each element of the integral image is calculated as the sum of intensities of all points above and to the left of the point in the image. The total intensity of any sub-window in an image can then be derived by subtracting the integral image value for the top left point of the sub-window from the integral image value for the bottom right point of the sub-window. Also intensities for adjacent sub-windows can be efficiently compared using particular combinations of integral image values from points of the sub-windows.

In Violla-Jones, a chain (cascade) of 32 classifiers based on rectangular (and increasingly refined) Haar features are used with the integral image by applying the classifiers to a sub-window within the integral image. For a complete analysis of an acquired image this sub-window is shifted incrementally across the integral image until the entire image has been covered.

In addition to moving the sub-window across the entire integral image, the sub window must also be scaled up/down to cover the possible range of face sizes. In Violla-Jones, a scaling factor of 1.25 is used and, typically, a range of about 10-12 different scales are required to cover the possible face sizes in an XVGA size image.

It will therefore be seen that the resolution of the integral image is determined by the smallest sized classifier sub-window, i.e. the smallest size face to be detected, as larger sized sub-windows can use intermediate points within the integral image for their calculations.

A number of variants of the original Violla-Jones algorithm are known in the literature. These generally employ rectangular, Haar feature classifiers and use the integral image techniques of Violla-Jones.

Even though Violla-Jones is significantly faster than other face detectors, it still requires significant computation and, on a Pentium class computer can just about achieve real-time performance. In a resource-restricted embedded system, such as hand held image acquisition devices (examples include digital cameras, hand-held computers or cellular phones equipped with cameras), it is not practical to run such a face detector at real-time frame rates for video. From tests within a typical digital camera, it is only possible to achieve complete coverage of all 10-12 sub-window scales with a 3-4 classifier cascade. This allows some level of initial face detection to be achieved, but with unacceptably high false positive rates.

US 2005/0147278, Rui et al describes a system for automatic detection and tracking of multiple individuals using multiple cues. Rui discloses using Violla-Jones as a fast face detector. However, in order to avoid the processing overhead of Violla-Jones, Rui instead discloses using an auto-initialization module which uses a combination of motion, audio and fast face detection to detect new faces in the frame of a video sequence. The remainder of the system employs well-known face tracking methods to follow existing or newly discovered candidate face regions from

frame to frame. It is also noted that Rui requires that some video frames be dropped in order to run a complete face detection.

DISCLOSURE OF THE INVENTION

According to a first aspect of the invention there is provided a method according to claim 1.

The first aspect of the present invention avoids the need for calculation of a complete highest resolution integral image for every acquired image in an image stream and so minimizes the integral image calculations which are required in face tracking systems. This either minimizes processing overhead for face detection and tracking or allows longer classifier chains to be employed during the frame-to-frame processing interval so providing higher quality results. This can significantly improve the performance and/or accuracy of real-time face detection and tracking.

In the preferred embodiment, when the invention is implemented in an image acquisition device during face detection, a subsampled copy of the acquired image is extracted from the camera hardware image acquisition subsystem and the integral image is calculated for this subsampled image. During face tracking, the integral image is only calculated for an image patch surrounding each candidate region.

In such an implementation, the process of face detection is spread across multiple frames. This approach is advantagous for effective implementation. In one example, digital image acquisition hardware is designed to subsample only to a single size. This invention takes advantage of the fact that when composing a picture, a face will typically be present for multiple frames of the video sequence. While this invention offers a significant improvement on the efficiency, the reduction in computation does not impact significantly on the initial detection of faces.

In the preferred embodiment, the 3-4 smallest sizes (lowest resolution) of subsampled images are used in cycle. In some cases, such as when the focus of the camera is set to infinity,

larger image subsamples may be included in the cycle as smaller (distant) faces may occur within the acquired image(s). In yet another embodiment, the number of subsampled images may change based on the estimated potential face sizes based on the estimated distance to the subject. Such distance may be estimated based on the focal length and focus distance, these acquisition parameters being available from other subsystems within the imaging appliance firmware.

By varying the resolution/scale of the sub-sampled image which is in turn used to produce the integral image, a single fixed size of classifier can be applied to the different sizes of integral image. Such an approach is particularly amenable to hardware embodiments where the subsampled image memory space can be scanned by a fixed size direct memory access (DMA) window and digital logic to implement a Haar-feature classifier chain can be applied to this DMA window. However, it will be seen that several sizes of classifier (in a software embodiment), or multiple fixed-size classifiers (in a hardware embodiment) could also be used.

A key advantage of this aspect of the invention is that from frame to frame only the lowest resolution integral image needs to be calculated.

Preferably, a full resolution image patch surrounding each candidate face region is acquired prior to the acquisition of the next image frame. An integral image is then calculated for each such image patch and a multi-scaled face detector is applied to each such image patch. Regions which are found by the multi-scaled face detector to be face regions are referred to as confirmed face regions.

The first aspect of the present invention avoids the need for motion and audio queues described in Rui and allows significantly more robust face detection and tracking to be achieved in a digital camera.

In a second aspect of the present invention, there is provided a method as claimed in claim 31.

According to this aspect, when face tracking detects a face region from a stream of images, the acquisition device firmware runs a face recognition algorithm at the location of the face using a database preferably stored on the acquisition device comprising personal identifiers

and their associated face parameters.

This aspect of the present invention mitigates the problems of algorithms using a single image for face detection and recognition which have lower probability of performing correctly.

In a third aspect of the invention, there is provided a method according to claim 36.

According to this aspect, the acquisition device includes an orientation sensor which indicates the likely orientation of faces in acquired images. The determined camera orientation is fed to face detection processes which then need only apply face detection for the likely orientation of faces. This improves processing requirements and/or face detection accuracy.

In a fourth aspect of the invention, there is provided a method according to claim 41.

This aspect of the invention improves the performance of a face tracking module by employing a motion sensor subsystem to indicate to the face tracking module, large motions of the acquisition device during a face tracking sequence.

Without such a sensor, where the acquisition device is suddenly moved by the user rather than slowly panned across a scene, candidate face regions in the next frame of a video sequence would be displaced beyond the immediate vicinity of the corresponding candidate region in the previous video frame and the face tracking module could fail to track the face requiring re-detection of the candidate.

In a fifth aspect of the invention there is provided a method according to claim 50.

By only running the face detector on regions predominantly including skin tones, more relaxed face detection can be used, as there is a higher chance that these skin-tone regions do in fact contain a face. So, faster face detection can be employed to more effectively provide similar quality results to running face detection over the whole image with stricter face detection required to positively detect a face.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the invention will now be described by way of example, with reference to the accompanying drawings, in which:

Figure 1 is a block diagram illustrating the principle components of an image processing

apparatus according to a preferred embodiment of the present invention;

Figure 2 is a flow diagram illustrating the operation of the image processing apparatus of Figure 1; and

Figures 3(a) to (d) shows examples of images processed by the apparatus of the preferred embodiment.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Fig 1 shows the primary subsystems of the face tracking system according to a preferred embodiment of the invention. The solid lines indicate the flow of image data; the dashed line indicate control inputs or information outputs (e.g. location(s) of detected faces) from a module. In this example an image processing apparatus can be a digital still camera (DSC), a video camera, a cell phone equipped with an image capturing mechanism or a hand help computer equipped with an internal or external camera.

A digital image is acquired in raw format from an image sensor (CCD or CMOS) [105] and an image subsampler [112] generates a smaller copy of the main image. Most digital cameras already contain dedicated hardware subsystems to perform image subsampling, for example to provide preview images to a camera display. Typically the subsampled image is provided in bitmap format (RGB or YCC). In the meantime the normal image acquisition chain performs post-processing on the raw image [110] which typically includes some luminance and color balancing. In certain digital imaging systems the subsampling may occur after such post-processing, or after certain post-processing filters are applied, but before the entire post-processing filter chain is completed.

The subsampled image is next passed to an integral image generator [115] which creates an integral image from the subsampled image. This integral image is next passed to a fixed size face detector [120]. The face detector is applied to the full integral image, but as this is an integral image of a subsampled copy of the main image, the processing required by the face detector is proportionately reduced. If the subsample is ¼ of the main image this implie the required processing time is only 25% of what would be required for the full image.

This approach is particularly amenable to hardware embodiments where the subsampled

image memory space can be scanned by a fixed size DMA window and digital logic to implement a Haar-feature classifier chain can be applied to this DMA window. However we do not preclude the use of several sizes of classifier (in a software embodiment), or the use of multiple fixed-size classifiers (in a hardware embodiment). The key advantage is that a smaller integral image is calculated.

After application of the fast face detector [280] any newly detected candidate face regions [141] are passed onto a face tracking module [111] where any face regions confirmed from previous analysis [145] are merged with the new candidate face regions prior to being provided [142] to a face tracker [290].

The face tracker [290] as will be explained later provides a set of confirmed candidate regions [143] back to the tracking module [111]. Additional image processing filters are applied by the tracking module [111] to confirm either that these confirmed regions [143] are face regions or to maintain regions as candidates if they have not been confirmed as such by the face tracker [290]. A final set of face regions [145] can be output by the module [111] for use elsewhere in the camera or to be stored within or in association with an acquired image for later processing either within the camera or offline; as well as to be used in the next iteration of face tracking.

After the main image acquisition chain is completed a full-size copy of the main image [130] will normally reside in the system memory [140] of the image acquisition system. This may be accessed by a candidate region extractor [125] component of the face tracker [290] which selects image patches based on candidate face region data [142] obtained from the face tracking module [111]. These image patches for each candidate region are passed to an integral image generator [115] which passes the resulting integral images to a variable sized detector [121] , as one possible example a VJ detector, which then applies a classifier chain, preferably at least a 32 classifier chain, to the integral image for each candidate region across a range of different scales.

The range of scales [144] employed by the face detector [121] is determined and supplied by the face tracking module [111] and is based partly on statistical information relating to the history of the current candidate face regions [142] and partly on external metadata determined from other subsystems within the image acquisition system.

As an example of the former, if a candidate face region has remained consistently at a particular size for a certain number of acquired image frames then the face detector [121] need only be applied at this particular scale and perhaps at one scale higher (i.e. 1.25 time larger) and one scale lower (i.e. 1.25 times lower).

As an example of the latter, if the focus of the image acquisition system has moved to infinity then it will be necessary to apply the smallest scalings in the face detector [121] Normally these scalings would not be employed as they must be applied a greater number of times to the candidate face region in order to cover it completely. It is worthwhile noting that the candidate face region will have a minimum size beyond which it not should decrease – this is in order to allow for localized movement of the camera by a user between frames. In some image acquisition systems which contain motion sensors it may be possible to track such localized movements and this information may be employed to further improved the selection of scales and the size of candidate regions.

The candidate region tracker [290] provides a set of confirmed face regions [143] based on full variable size face detection of the image patches to the face tracking module [111]. Clearly, some candidate regions will have been confirmed while others will have been rejected and these can be explicitly returned by the tracker [290] or can be calculated by the tracking module [111] by analyzing the difference between the confirmed regions [143] and the candidate regions [142]. In either case, the face tracking module [111] can then apply alternative tests to candidate regions rejected by the tracker [290] (as explained below) to determine whether these should be maintained as candidate regions [142] for the next cycle of tracking or whether these should indeed be removed from tracking.

Once the set of confirmed candidate regions [145] has been determined by the face tracking module [111], the module [111] communicates with the sub-sampler [112] to determine when the next acquired image is to be sub-sampled and so provided to the detector [280] and also to provide the resolution [146] at which the next acquired image is to be sub-sampled.

It will be seen that where the detector [280] does not run when the next image is acquired, the candidate regions [142] provided to the extractor [125] for the next acquired image will be the

regions [145] confirmed by the tracking module [111] from the last acquired image. On the other hand, when the face detector [280] provides a new set of candidate regions [141] to the face tracking module [111], these candidate regions are merged with the previous set of confirmed regions [145] to provide the set of candidate regions [142] to the extractor [125] for the next acquired image.

Fig 2 shows the main workflow in more detail. The process is split into (i) a detection/initialization phase which finds new candidate face regions [141] using the fast face detector [280] which operates on a subsampled version of the full image; (ii) a secondary face detection process [290] which operates on extracted image patches for the candidate regions [142], which are determined based on the location of faces in one or more previously acquired image frames and (iii) a main tracking process which computes and stores a statistical history of confirmed face regions [143]. Although we show the application of the fast face detector [280] occurring prior to the application of the candidate region tracker [290] the order is not critical and the fast detection is not necessarily executed on every frame or in certain circumstances may be spread across multiple frames.

Thus, in step 205 the main image is acquired and in step 210 primary image processing of that main image is performed as described in relation to Figure 1. The sub-sampled image is generated by the subsampler [112] and an integral image is generated therefrom by the generator [115], step 211 as described previously. The integral image is passed to the fixed size face detector [120] and the fixed size window provides a set of candidate face regions [141] within the integral image to the face tracking module, step 220. The size of these regions is determined by the sub-sampling scale [146] specified by the face tracking module to the sub-sampler and this scale is based on the analysis of the previous sub-sampled/integral images by the detector [280] and patches from previous acquired images by the tracker [290] as well as other inputs such as camera focus and movement.

The set of candidate regions [141] is merged with the existing set of confirmed regions [145] to produce a merged set of candidate regions [142] to be provided for confirmation, step 242.

For the candidate regions [142] specified by the face tracking module 111, the candidate

region extractor [125] extracts the corresponding full resolution patches from an acquired image, step 225. An integral image is generated for each extracted patch, step 230 and a variable sized face detection is applied by the face detector 121 to each such integral image patch, for example, a full Violla-Jones analysis. These results [143] are in turn fed back to the face-tracking module [111], step 240.

The tracking module [111] processes these regions [143] further before a set of confirmed regions [145] is output. In this regard, additional filters can be applied by the module 111 either for regions [143] confirmed by the tracker [290] or for retaining candidate regions [142] which may not have been confirmed by the tracker 290 or picked up by the detector [280], step 245.

For example, if a face region had been tracked over a sequence of acquired images and then lost, a skin prototype could be applied to the region by the module [111] to check if a subject facing the camera had just turned away. If so, this candidate region could be maintained for checking in the next acquired image to see if the subject turns back to face the camera.

Depending on the sizes of the confirmed regions being maintained at any given time and the history of their sizes, e.g. are they getting bigger or smaller, the module 111, determines the scale [146] for sub-sampling the next acquired image to be analysed by the detector [280] and provides this to the sub-sampler [112], step 250.

It will be seen that typically the fast face detector [280] need not run on every acquired image. So for example, where only a single source of sub-sampled images is available, if a camera acquires 60 frames per second, 15-25 sub-sampled frames per second (fps) may be required to be provided to the camera display for user previewing. Clearly, these images need to be sub-sampled at the same scale and at a high enough resolution for the display. Some or all of the remaining 35-45 fps can be sampled at the scale required by the tracking module [111] for face detection and tracking purposes.

The decision on the periodicity in which images are being selected from the stream may be based on a fixed number or alternatively be a run-time variable. In such cases, the decision on the next sampled image may be determined on the processing time it took for the previous image, in order to maintain synchronicity between the captured real-time stream and the face tracking processing. Thus in a complex image environment the sample rate may decrease.

Alternatively, the decision on the next sample may also be performed based on processing of the content of selected images. If there is no significant change in the image stream, the full face tracking process will not need to be performed. In such cases, although the sampling rate may be constant, the images will undergo a simple image comparison and only if it is decided that there is justifiable differences, will the face tracking algorithms be launched.

It will also be noted that the face detector [280] need not run at regular intervals. So for example, if the camera focus is changed significantly, then the face detector may need to run more frequently and particularly with differing scales of sub-sampled image to try to detecting faces which should be changing in size. Alternatively, where focus is changing rapidly, the detector [280] could be skipped for intervening frames, until focus has stabilised. However, it is generally only when focus goes to infinity that the highest resolution integral image must be produced by the generator [115].

In this latter case, the detector may not be able to cover the entire area of the acquired, subsampled, image in a single frame. Accordingly the detector may be applied across only a portion of the acquired, subsampled, image on a first frame, and across the remaining portion(s) of the image on subsequent acquired image frames. In a preferred embodiment the detector is applied to the outer regions of the acquired image on a first acquired image frame in order to catch small faces entering the image from its periphery, and on subsequent frames to more central regions of the image.

An alternative way of limiting the areas of an image to which the face detector 120 is to be applied comprises identifying areas of the image which include skin tones. US 6,661,907 discloses one such technique for detecting skin tones and subsequently only applying face detection in regions having a predominant skin colour.

In one embodiment of the present invention, skin segmentation 190 is preferably applied to the sub-sampled version of the acquired image. If the resolution of the sub-sampled version is not sufficient, then a previous image stored image store 150 or a next sub-sampled image can be used as long as the two image are not too different in content from the current acquired image. Alternatively, skin segmentation 190 can be applied to the full size video image 130.

In any case, regions containing skin tones are identified by bounding rectangles and these bounding rectangles are provided to the integral image generator 115 which produces integral image patches corresponding to the rectangles in a manner similar to the tracker integral image generator 115.

Not alone does this approach reduce the processing overhead associated with producing the integral image and running face detection, but in the present embodiment, it also allows the face detector 120 to apply more relaxed face detection to the bounding rectangles, as there is a higher chance that these skin-tone regions do in fact contain a face. So for a VJ detector 120, a shorter classifier chain can be employed to more effectively provide similar quality results to running face detection over the whole image with longer VJ classifiers required to positively detect a face.

Further improvements to face detection are also possible. For example, it has been found that face detection is very dependent on illumination conditions and so small variations in illumination can cause face detection to fail, causing somewhat unstable detection behavior.

In present embodiment, confirmed face regions 145 are used to identity regions of a subsequently acquired subsampled image on which luminance correction should be performed to bring the regions of interest of the image to be analyzed to the desired parameters. One example of such correction is to improve the luminance contrast within the regions of the subsampled image defined by the confirmed face regions 145.

Contrast enhancement is well-known and is typically used to increased the local contrast of an image, especially when the usable data of the image is represented by close contrast values. Through this adjustment, the intensities for pixels of a region when represented on a histogram which would otherwise be closely distributed can be better distributed. This allows for areas of lower local contrast to gain a higher contrast without affecting the global contrast. Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values.

The method is useful in images with backgrounds and foregrounds that are both bright or both dark. In particular, the method can lead to better detail in photographs that are over or under-exposed.

Alternatively, this luminance correction could be included in the computation of an

"adjusted" integral image in the generators 115.

In another improvement, when face detection is being used, the camera application is set to dynamically modify the exposure from the computed default to a higher values (from frame to frame, slightly overexposing the scene) until the face detection provides a lock onto a face.

In a separate embodiment, the face detector 120 will be applied only to the regions that are substantively different between images. Note that prior to comparing two sampled images for change in content, a stage of registration between the images may be needed to remove the variability of changes in camera, caused by camera movement such as zoom, pan and tilt.

It will be seen that it is possible to obtain zoom information from camera firmware and it is also possible using software techniques which analyze images in camera memory 140 or image store 150 to determine the degree of pan or tilt of the camera from one image to another.

However, in one embodiment of the invention, the acquisition device is provided with a motion sensor 180, Figure 1, to determine the degree and direction of pan from one image to another so avoiding the processing requirement of determining camera movement in software.

Many digital cameras have begun to incorporate such motion sensors – normally based on accelerometers, but optionally based on gyroscopic principals - within the camera, primarily for the purposes of warning or compensating for hand shake during main image capture. US 4,448,510, Murakoshi discloses such a system for a conventional camera, or US 6,747,690, Molgaard discloses accelerometer sensors applied within a modern digital camera.

Where a motion sensor is incorporated in a camera it will typically be optimized for small movements around the optical axis. A typical accelerometer incorporates a sensing module which generates a signal based on the acceleration experienced and an amplifier module which determines the range of accelerations which can effectively be measured. Modern accelerometers allow software control of the amplifier stage which allows the sensitivity to be adjusted.

The motion sensor 180 could equally be implemented with MEMS sensors of the sort which will be incorporated in next generation consumer cameras and camera-phones.

In any case, when the camera is operable in face tracking mode, i.e. constant video acquisition as distinct from acquiring a main image, shake compensation is typically not used

because image quality is lower. This provides the opportunity to configure the motion sensor 180, to sense large movements, by setting the motion sensor amplifier module to low gain. The size and direction of movement detected by the sensor 180 is provided to the face tracker 111. The approximate size of faces being tracked is already known and this enables an estimate of the distance of each face from the camera. Accordingly, knowing the approximate size of the large movement from the sensor 180 allows the approximate displacement of each candidate face region to be determined, even if they are at differing distances from the camera.

Thus, when a large movement is detected, the face tracker 111 shifts the location of candidate regions as a function of the direction and size of the movement. Alternatively, the size of the region over which the tracking algorithms are applied may also be enlarged (and, if necessary, the sophistication of the tracker may be decreased to compensate for scanning a larger image area) as a function of the direction and size of the movement.

When the camera is actuated to capture a main image, or when it exits face tracking mode for any other reason, the amplifier gain of the motion sensor 180 is returned to normal, allowing the main image acquisition chain 105,110 for full-sized images to employ normal shake compensation algorithms based on information from the motion sensor 180.

In alternative embodiments, sub-sampled preview images for the camera display can be fed through a separate pipe than the images being fed to and supplied from the image sub-sampler [112] and so every acquired image and its sub-sampled copies can be available both to the detector [280] as well as for camera display.

In addition to periodically acquiring samples from a video stream, the process may also be applied to a single still image acquired by a digital camera. In this case, the stream for the face tracking comprises a stream of preview images and the final image in the series is the full resolution acquired image. In such a case, the face tracking information can be verified for the final image in a similar fashion to that described in Figure 2. In addition, the information such as coordinates or mask of the face may be stored with the final image. Such data for example may fit as an entry in the saved image header, for future post processing, whether in the acquisition device or at a later stage by an external device.

Turning now to Figure 3 which illustrates the operation of the preferred embodiment through a worked example. ***Fig 3:*** (a) illustrates the result at the end of a detection & tracking cycle on a frame of video; two confirmed face regions [301, 302] of different scales are shown. In the present embodiment, for pragmatic reasons, each face region has a rectangular bounding box; as it is easier to make computations on rectangular regions. This information is recorded and output as [145] by the tracking module [111] of fig 1.

Based on the history of the face regions [301,302], the tracking module [111] decides to run fast face tracking with a classifier window of the size of face region [301] with an integral image being provided and analyzed accordingly.

Fig 3(b) shows the situation after the next frame in a video sequence is captured and the fast face detector has been applied to the new image. Both faces have moved [311, 312] and are shown relative to the previous face regions [301, 302]. A third face region [303] has appeared and has been detected by the fast face detector [303]. In addition the fast face detector has found the smaller of the two previously confirmed faces [304] because it is at the correct scale for the fast face detector. Regions [303] and [304] are supplied as candidate regions [141] to the tracking module [111]. The tracking module merges this new candidate region information [141], with the previous confirmed region information [145] comprising regions [301] [302] to provide a set of candidate regions comprising regions [303],[304] and [302] to the candidate region extractor [290]. The tracking module [111] knows that the region [302] has not been picked up by the detector [280]. This may be because the face has in either disappeared, remains at a size that could not have been detected by the detector [280] or has changed size to a size that could not have been detected by the detector [280]. Thus, for this region, the module [111] will specify a large patch [305], Fig 3(c) around the region [302] to be checked by the tracker [290]. Only the region [303] bounding the newly detected face candidate needs to be checked by the tracker [290], whereas because the face [301] is moving a relatively large patch [306] surrounding this region is specified to the tracker [290].

Fig 3(c) shows the situation after the candidate region extractor operates upon the image; candidate regions [306, 305] around both of the confirmed face regions [301, 302] from the

previous video frame as well as new region [303] are extracted from the full resolution image [130]; the size of these candidate regions having been calculated by the face tracking module [111] based partly on partly on statistical information relating to the history of the current face candidate and partly on external metadata determined from other subsystems within the image acquisition system. These extracted candidate regions are now passed on to the variable sized face detector [121] which applies a VJ face detector to the candidate region over a range of scales; the locations of any confirmed face regions are then passed back to the face tracking module [111].

Fig 3(d) shows the situation after the face tracking module [111] has merged the results from both the fast face detector [280] and the face tracker [290] and applied various confirmation filters to the confirmed face regions. Three confirmed face regions have been detected [307, 308, 309] within the patches [305,306,303]. The largest region [307] was known but had moved from the previous video frame and relevant data is added to the history of that face region;. The other previously known region [308] which had moved was also detected by the fast face detector which serves as a double-confirmation and these data are added to its history. Finally a new face region [303] was detected and confirmed and a new face region history must be initiated for this newly detected face. These three face regions are used to provide a set of confirmed face regions [145] for the next cycle.

It will be seen that there are many possible applications for the regions 145 supplied by the face tracking module. For example, the bounding boxes for each of the regions [145] can be superimposed on the camera display to indicate that the camera is automatically tracking detected face(s) in a scene. This can be used for improving various pre-capture parameters. One example is exposure, ensuring that the faces are well exposed. Another example is auto-focusing, by ensuring that focus is set on a detected face or indeed to adjust other capture settings for the optimal representation of the face in an image.

The corrections may be done as part of the pre-processing adjustments. The location of the face tracking may also be used for post processing and in particular selective post processing where the regions with the faces may be enhanced. Such examples include sharpening, enhancing

saturation, brightening or increasing local contrast. The preprocessing using the location of faces may also be used on the regions without the face to reduce their visual importance, for example through selective blurring, desaturation, or darkening.

Where several face regions are being tracked, then the longest lived or largest face can be used for focusing and can be highlighted as such. Also, the regions [145] can be used to limit the areas on which for example red-eye processing is performed when required.

Other post-processing which can be used in conjunction with the light-weight face detection described above is face recognition. In particular, such an approach can be useful when combined with more robust face detection and recognition either running on the same or an off-line device that has sufficient resources to run more resource consuming algorithms

In this case, the face tracking module [111] reports the location of any confirmed face regions [145] to the in-camera firmware, preferably together with a confidence factor.

When the confidence factor is sufficiently high for a region, indicating that at least one face is in fact present in an image frame, the camera firmware runs a light-weight face recognition algorithm [160] at the location of the face, for example a DCT-based algorithm. The face recognition algorithm [160] uses a database [161] preferably stored on the camera comprising personal identifiers and their associated face parameters.

In operation, the module [160] collects identifiers over a series of frames. When the identifiers of a detected face tracked over a number of preview frames are predominantly of one particular person, that person is deemed by the recognition module to be present in the image. The identifier of the person, and the last known location of the face, is stored either in the image (in a header) or in a separate file stored on the camera storage [150]. This storing of the person's ID can occur even when the recognition module [160[failed for the immediately previous number of frames but for which a face region was still detected and tracked by the module [111].

When the image is copied from camera storage to a display or permanent storage device such as a PC (not shown), the person ID's are copied along with the images. Such devices are generally more capable of running a more robust face detection and recognition algorithm and

then combining the results with the recognition results from the camera, giving more weight to recognition results from the robust face recognition (if any). The combined identification results are presented to the user, or if identification was not possible, the user is asked to enter the name of the person that was found. When the user rejects an identification or a new name is entered, the PC retrains its face print database and downloads the appropriate changes to the capture device for storage in the light-weight database [161].

It will be seen that when multiple confirmed face regions [145] are detected, the recognition module [160] can detect and recognize multiple persons in the image.

It is possible to introduce a mode in the camera that does not take a shot until persons are recognized or until it is clear that persons are not present in the face print database, or alternatively displays an appropriate indicator when the persons have been recognized. This would allow reliable identification of persons in the image.

This aspect of the present system solves the problem where algorithms using a single image for face detection and recognition may have lower probability of performing correctly. In one example, for recognition, if the face is not aligned within certain strict limits it is not possible to accurately recognize a person. This method uses a series of preview frames for this purpose as it can be expected that a reliable face recognition can be done when many more variations of slightly different samples are available.

Further improvements to the efficiency of the system described above are possible. For example, conventional face detection algorithms typically employ methods or use classifiers to detect faces in a picture at different orientations: 0, 90, 180 and 270 degrees.

According to a further aspect of the present invention, the camera is equipped with an orientation sensor 170, Figure 1. This can comprise a hardware sensor for determining whether the camera is being held upright, inverted or tilted clockwise or anti-clockwise. Alternatively, the orientation sensor can comprise an image analysis module connected either to the image acquisition hardware 105, 110 or camera memory 140 or image store 150 for quickly determining whether images are being acquired in portrait or landscape mode and whether the camera is tilted

clockwise or anti-clockwise.

Once this determination is made, the camera orientation can be fed to one or both of the face detectors 120, 121. The detectors need then only apply face detection according to the likely orientation of faces in an image acquired with the determined camera orientation. This aspect of the invention can either significantly reduce the face detection processing overhead, for example, by avoiding the need to employ classifiers which are unlikely to detect faces or increase its accuracy by running classifiers more likely to detects faces in a given orientation more often.

Claims:

1. A method of tracking faces in an image stream using a digital image acquisition device comprising:

a. receiving a new acquired image from said image stream, said acquired image potentially including one or more face regions;

b. sub-sampling said acquired image at a specified resolution to provide a first-sub-sampled image;

c. calculating for a least a portion of said sub-sampled acquired image a corresponding integral image;

d. applying at least a fixed size face detection to at least a portion of said integral image to provide a set of candidate face regions, each candidate face region having a given size and a respective location;

e. responsive to the size and location of set of candidate face regions and any previously detected face regions, adjusting the resolution at which a next acquired image is sub-sampled; and

f. repeating steps a. to e.

2. A method as claimed in claim 1 further comprising:

tracking candidate face regions of different sizes from a plurality of images of said image stream.

3. A method as claimed in claim 1 further comprising the step of:

merging said set of candidate face regions with any previously detected face regions to provide a set of candidate face regions of potentially different sizes.

4. A method as claimed in claim 3 further comprising the step of:

for each region of said acquired image corresponding to a region of said merged set of candidate face regions:

calculating an integral image; and

applying variable sized face detection to each merged region integral image to provide a set of confirmed face regions and a set of rejected face regions.

5. A method as claimed in claim 4 further comprising the step of:

checking any rejected face regions based on alternative criteria from said fixed and variable sized face detection; and

responsive to said checking indicating any of said rejected face regions is a face region, adding said previously rejected face region to said set of confirmed face regions.

6. A method as claimed in claim 4 wherein said merging comprises merging a set of candidate face regions for an acquired image with said set of confirmed face regions for a previously acquired image.

7. A method as claimed in claim 1 wherein said adjusting comprises cycling through a set of approximately 4 sub-sampling resolutions.

8. A method as claimed in claim 1 wherein said adjusting is responsive to said image being acquired at infinite focus for adjusting said sub-sampling resolution to maximum resolution.

9. A method as claimed in claim 1 wherein said steps a. to e. are performed periodically on a selected plurality of images of an image stream, said plurality of images including a main acquired image chronologically following a plurality of images of said selected plurality of images.

10. A method as claimed in claim 9 where the time interval between said selected images is fixed.

11. A method as claimed in claim 10 where the time interval between said selected images is variable and determined during execution.

12. A method as claim in claim 11 wherein said time interval is a function of the time taken to process a previous main acquired image

13. A method as claimed in claim 12 wherein said time interval is dependent in changes in the image content of one or more of said selected plurality of images.

14. A method as claimed in claim 1 wherein said repeating steps a. to e. is performed responsive to changes in content from image to image in said image stream.

15. A method as claimed in claim 4 wherein said regions of said acquired image corresponding to a region of said merged set of candidate face regions comprise regions surrounding respective regions of said merged set of candidate face regions.

16. A method as claimed in claim 1 wherein said applying fixed size face detection comprises applying a cascade of Haar classifiers of a fixed size to said integral image.

17. A method as claimed in claim 12 wherein said cacade comprises 32 classifiers.

18. A method as claimed in claim 4 wherein applying variable sized face detection comprises applying a plurality of cascades of Haar classifiers of a varying size to each merged region integral image.

19. A method as claimed in claim 18 wherein up to 10 to 12 sizes of cascades of Haar classifiers are applied and preferably between 3 and 4 sizes of Haar classifiers are applied.

20. A method as claimed in claim 19 wherein each cacade comprises 32 classifiers.

21. A method as claimed in claim 5 wherein said checking comprises applying a skin prototype to any rejected face region.

22. A method as claimed in claim 1 further comprising the step of:
displaying an acquired image; and
superimposing said tracked candidate face regions on said displayed acquired image.

23. A method as claimed in claim 9 further comprising storing at least one of the size and location of at least some of said set of candidate face regions in association with said main acquired image.

24. A method as claimed in claim 1 further comprising the step of:
responsive to said acquired image being captured with a flash, analysing regions of said acquired image corresponding to said tracked candidate face regions for red-eye defects.

25. A method as claimed in claim 23 further comprising the step of:
performing spatially selective post processing of said main acquired image based on said stored candidate face regions' size or location.

26. A method as claimed in claim 25 wherein said selective image processing comprises color correction, sharpening, blurring, saturation, subsampling, compression, or a combination thereof.

27. A method as claimed in claim 24 further comprising the step of:
correcting in said acquired image any red-eye defects.

28. A method as claimed in claim 24 further comprising the step of:
storing with said acquired image an indication of any red-eye defects.

29. An image processing apparatus for tracking faces in an image stream arranged to iteratively:

a. receive a new acquired image from said image stream, said image potentially including one or more face regions;

b. sub-sample said acquired image at a specified resolution to provide a sub-sampled image;

c. calculate for a least a portion of said sub-sampled acquired image a corresponding integral image;

d. apply at least a fixed size face detection to at least a portion of said integral image to provide a set of candidate face regions; and

e. responsive to said set of candidate face regions and any previously detected candidate face regions, adjust the resolution at which a next acquired image is sub-sampled.

30. An apparatus according to claim 29 comprising one of a digital still camera, a video camera, a cell phone equipped with an image capturing mechanism or a hand help computer equipped with an internal or external camera

31. A method of recognizing faces in an image stream using a digital image acquisition device comprising:

a. providing a database comprising an identifier and associated parameters for each of a number of faces to be recognized;

b. receiving a new acquired image from said image stream, said acquired image potentially including one or more face regions;

c. sub-sampling said acquired image at a specified resolution to provide a first-sub-

sampled image;

d. calculating for a least a portion of said sub-sampled acquired image a corresponding integral image;

e. applying face detection to at least a portion of said integral image to provide a set of candidate face regions, each candidate face region having a given size and a respective location;

f. selectively applying face recognition using said database to at least some of said candidate face regions to provide an identifier for any face recognized in a candidate face region;

g. storing an identifier for any recognized face in association with at least one image of said image stream; and

g. repeating steps b. to g.

32. A method as claimed in claim 31 comprising

responsive to the size and location of set of candidate face regions and any previously detected face regions, adjusting the resolution at which a next acquired image is sub-sampled.

33. A method as claimed in claim 31 wherein said face detection provides a level of confidence for each candidate region and wherein face recognition is applied to candidate face regions having a confidence level higher than a predetermined threshold.

34. A method as claimed in claim 31 wherein said identifiers are stored in association with a full-resolution image of said image stream.

35. An image processing apparatus for recognizing faces in an image stream, said apparatus including a database comprising an identifier and associated parameters for each of a number of faces to be recognized, and being arranged to iteratively:

b. receive a new acquired image from said image stream, said acquired image potentially including one or more face regions;

c. sub-sample said acquired image at a specified resolution to provide a first-sub-sampled image;

d. calculate for a least a portion of said sub-sampled acquired image a corresponding integral image;

e. apply face detection to at least a portion of said integral image to provide a set of candidate face regions, each candidate face region having a given size and a respective location;

f. selectively apply face recognition using said database to at least some of said candidate face regions to provide an identifier for any face recognized in a candidate face region; and

g. store an identifier for any recognized face in association with at least one image of said image stream.

36. A method of detecting faces in an image stream using a digital image acquisition device comprising:

a. determining an orientation of said acquisition device for at least one image of said image stream

b. acquiring an image from said image stream, said acquired image potentially including one or more face regions; and

c. applying face detection to at least a portion of said acquired image to provide a set of candidate face regions according to said determined orientation, each candidate face region having a given size and a respective location.

37. A method according to claim 36 wherein said face detection selectively employs one sequence of classifiers from an available plurlality of classifiers sequences, each classifier sequence being at a given orientation.

38. An image processing apparatus for detecting faces in an image stream, said apparatus arranged to:

a. determine an orientation of said acquisition device for at least one image of said image stream

b. acquire an image from said image stream, said acquired image potentially including one or more face regions; and

c. apply face detection to at least a portion of said acquired image to provide a set of candidate face regions according to said determined orientation, each candidate face region having a given size and a respective location.

39. An image processing apparatus according to claim 38 comprising a hardware orientation sensor for indicating the orientation of said apparatus.

40. An image processing apparatus according to claim 38 in which said apparatus is arranged to analyze the content of one or more images of said image stream to determine the orientation of said apparatus.

41. A method of tracking faces in an image stream using a digital image acquisition device comprising:

a. receiving a new acquired image from said image stream, said acquired image potentially including one or more face regions;

b. receiving an indication of relative movement of said new acquired image relative to a previously acquired image, said previously acquired image having an associated set of candidate face regions, each candidate face region having a given size and a respective location;

c. applying adjusted face detection to at least a portion of said new acquired image in the vicinity of said candidate face regions as a function of said movement, to provide an updated set of candidate face regions; and

e. repeating steps a. to c.

42. A method as claimed in claim 41 wherein said indication of relative movement comprises a size and direction of movement.

43. A method as claimed in claim 41 wherein said adjusted face detection comprises: prior to applying face detection, shifting said associated set of candidate face regions as a function of said movement;

44. A method as claimed in claim 43 wherein said face regions are shifted as a function of their size and as a function of said movement.

45. A method as claimed in claim 41 wherein said adjusted face detection comprises: applying face detection to a region of said new acquired image comprising the candidate regions associated with the previously acquired image expanded as a function of said movement.

46. A method as claimed in claim 45 wherein said regions are expanded as a function of their original size and as a function of said movement.

47. An image processing apparatus for tracking faces in an image stream, arranged to iteratively:

a. receive a new acquired image from said image stream, said new acquired image potentially including one or more face regions;

b. receive an indication of relative movement of said new acquired image relative to a previously acquired image, said previously acquired image having an associated set of candidate face regions, each candidate face region having a given size and a respective location; and

c. apply adjusted face detection to at least a portion of said new acquired image in the vicinity of said candidate face regions as a function of said movement, to provide an

updated set of candidate face regions.

48. An image processing apparatus as claimed in claim 47 comprising a motion sensor, said motion sensor comprising an accelerometer and a controlled gain amplifier connected to said accelerometer, said apparatus being arranged to set the gain of said amplifier relatively low for acquisition of a high resolution image and to set the gain of said amplifier relatively high during acquisition of a stream of relatively low resolution images.

49. An image processing apparatus as claimed in claim 47 including a motion sensor, said motion sensor comprising a MEMS sensor.

50. A method of detecting faces in an image stream using a digital image acquisition device comprising:

a. receiving a new acquired image from said image stream, said acquired image potentially including one or more face regions;

b. sub-sampling said acquired image at a specified resolution to provide a sub-sampled image;

c. identifying one or more regions of said acquired image predominantly including skin tones;

d. calculating for a least each of said skin tone regions of said sub-sampled acquired image a corresponding integral image;

e. applying face detection to at least a portion of said integral images to provide a set of candidate face regions, each candidate face region having a given size and a respective location; and

f. repeating steps a. to e.

51. A method as claimed in claim 50 wherein said identifying is performed on said sub-sampled image.

52. A method as claimed in claim 50 wherein said face detection is performed with relaxed face detection parameters.

53. A method as claimed in claim 50 further comprising:
for any of a set of candidate face regions associated with a previously acquired image, enhancing the contrast of the luminance characteristics of the corresponding regions of said new acquired image.

54. A method as claimed in claim 53 wherein said enhancing is performed on said sub-sampled image.

55. A method as claimed in claim 53 wherein said enhancing is performed during calculation of said integral image.

56. A method as claimed in claim 50 where in a face detection mode of said digital image acquisition device, each new acquired image is acquired with progressively increased exposure parameters until at least one candidate face region is detected.

57. An image processing apparatus for detecting faces in an image stream, arranged to iteratively:
a. receive a new acquired image from said image stream, said acquired image potentially including one or more face regions;
b. sub-sample said acquired image at a specified resolution to provide a sub-sampled image;

c. identify one or more regions of said acquired image predominantly including skin tones;

d. calculate for a least each of said skin tone regions of said sub-sampled acquired image a corresponding integral image; and

e. apply face detection to at least a portion of said integral images to provide a set of candidate face regions, each candidate face region having a given size and a respective location.

1/3

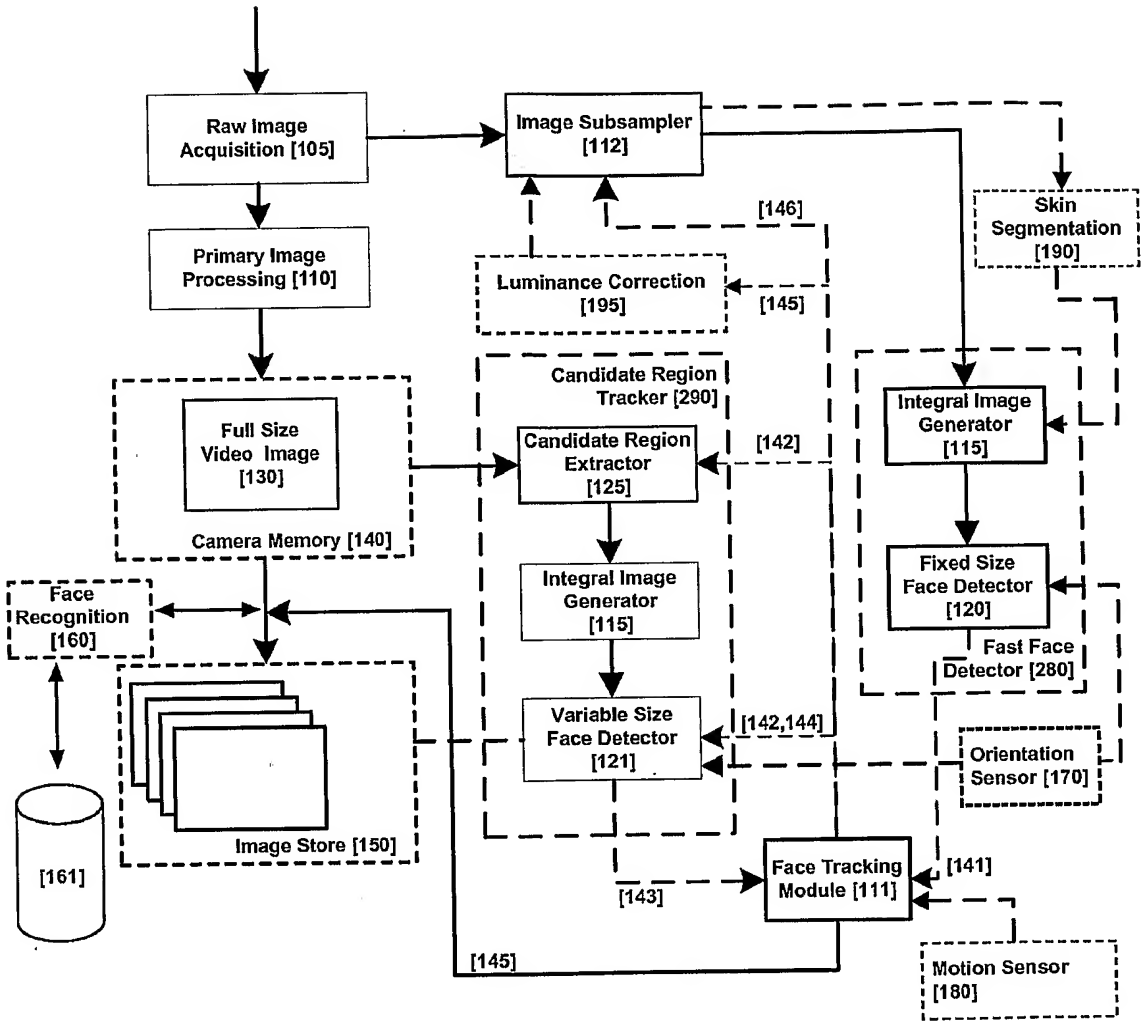

Raw Image Acquisition [105]
Image Subsampler [112]
Skin Segmentation [190]
[146]
Primary Image Processing [110]
Luminance Correction [195]
[145]
Candidate Region Tracker [290]
Integral Image Generator [115]
Full Size Video Image [130]
Candidate Region Extractor [125]
[142]
Camera Memory [140]
Integral Image Generator [115]
Fixed Size Face Detector [120]
Face Recognition [160]
Fast Face Detector [280]
Variable Size Face Detector [121]
[142,144]
Orientation Sensor [170]
Image Store [150]
[161]
Face Tracking Module [111]
[141]
[143]
[145]
Motion Sensor [180]

Figure 1

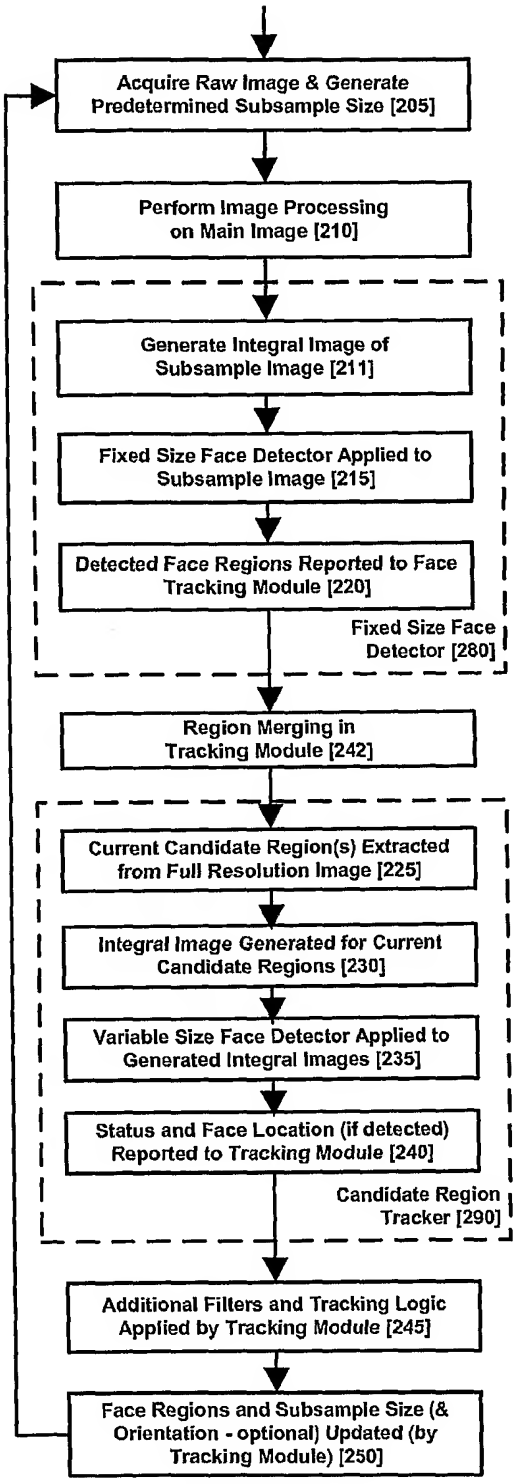


Figure 2

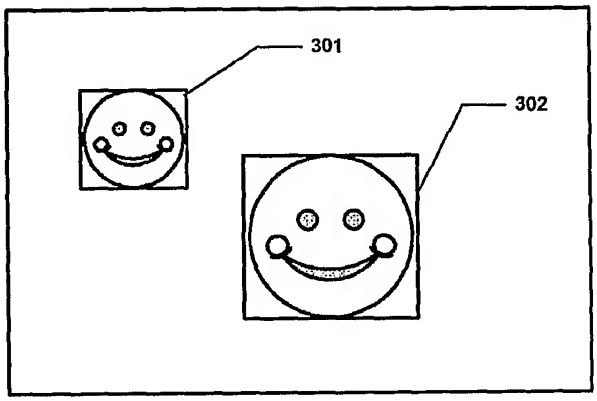


Figure 3(a)

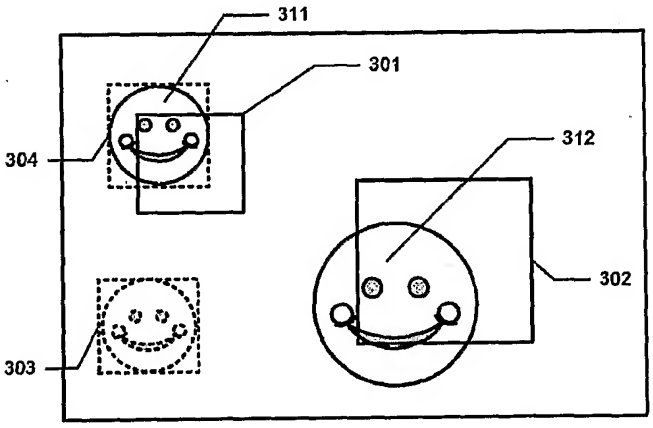


Figure 3(b)

3/3

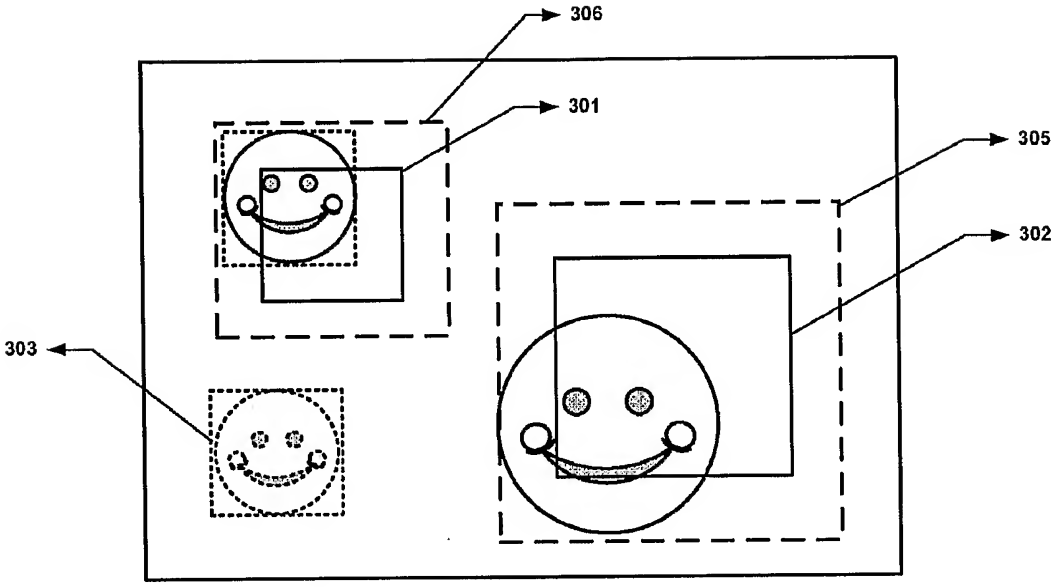


Figure 3(c)

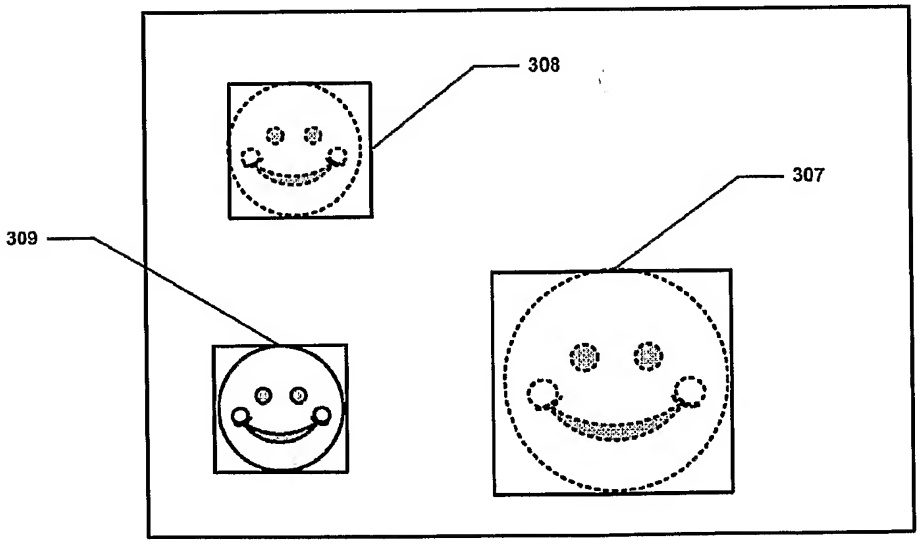


Figure 3(d)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US06/32959

A. CLASSIFICATION OF SUBJECT MATTER

IPC: **G06K 9/00**(2006.01),**9/32**(2006.01),**9/34**(2006.01),**9/40**(2006.01)

USPC: 382/115,118,173,254,299,300

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 382/115,118,173,254,299,300

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 7,082,212 A (LIU et al) 25 July 2006 (25.07.2006), column 10, lines 12-53.	1-57

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 January 2007 (30.01.2007)	06 MAR 2007
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (571) 273-3201	Authorized officer Amir Alavi Telephone No. 571-272-7386

Form PCT/ISA/210 (second sheet) (April 2005)